非凸制約付き最小化問題の弱劣モジュラ・弱優モジュラ構造 (高度情報化社会に向けた数理最適化の新潮流)

坂上, 晋作

坂上, 晋作. 非凸制約付き最小化問題の弱劣モジュラ・弱優モジュラ構造 (高度情報化社会に向けた数理最適化の新潮流). 数理解析研究所講究録 2019, 2108: 160-167

数理解析研究所講究録 第2108巻 2019年 160-167

非凸制約付き最小化問題の弱劣モジュラ・弱優モジュラ構造

NTT コミュニケーション科学基礎研究所　坂上晋作
Shinsaku Sakaue
NTT Communication Science Laboratories
sakaue.shinsaku@lab.ntt.co.jp

概 要

非凸制約付き最小化問題は，機械学習などの分野で様々な応用を持つ重要な問題である．この問題に対する理論保証を得るための手段として，制限強凸性と制限平滑性を用いた解析が注目されている．特に，最近の研究結果により，制限強凸性と制限平滑性を持つ最小化問題は，弱劣モジュラ関数最大化問題と深い関わりを持つことが明らかになっており，集合関数最大化問題の観点から非凸制約付き最小化問題を解析するアプローチが有望視されている．本項では，制限強凸性・制限平滑性が弱劣モジュラ性のみならず弱優モジュラ性とも関係を持つことを示し，この事実を用いて，ℓ_0-制約付き最小化問題に対する乱択 FPT 近似アルゴリズムを与える．また，さらに複雑な制約を持つ問題として，非ゼロ添字についての単調劣モジュラ関数の値が制限された最小化問題を考え，この問題に対する貪欲法や射影勾配法の理論保証を与える．

1　導入

解の非ゼロ成分の添字集合についての制約を持つ最適化問題は，様々な場面で現れる [1, 7]．そのような問題は，しばしば以下のような非凸制約付き最小化問題として記述される [10, 11, 12]：

$$\underset{\mathbf{x}\in\mathbb{R}^{[d]}}{\text{minimize}}\ \ l(\mathbf{x}) \quad \text{subject to}\ \ \text{supp}(\mathbf{x}) \in \mathcal{F}. \tag{1}$$

ただし，$[d] \coloneqq \{1, \ldots, d\}$ は添字の台集合であり，$\text{supp}(\mathbf{x}) \subseteq [d]$ は $\mathbf{x}$ の非ゼロ成分の添字集合である（以下，$\mathbf{x}$ のサポートと呼ぶ）．$\mathcal{F} \subseteq 2^{[d]}$ は解として許容されるサポートの集合である．このように記述される制約を満たす実行可能領域は，一般に非凸集合となる．$l : \mathbb{R}^{[d]} \to \mathbb{R}$ は連続的微分可能な目的関数であり，後述の制限強凸性と制限平滑性を満たす．問題 (1) に対してしばしば用いられる最適化手法として，射影勾配法に基づくものがある．機械学習におけるスパース最適化の分野では，このような手法は iterative hard thresholding (IHT) [3, 13] や hard thresholding pursuit (HTP) [9, 20] という名で広く研究されている．一方で，Elenberg らの最近の研究 [8] によって，問題 (1) から導出される集合関数

$$F(\mathsf{S}) \coloneqq l(0) - \min_{\text{supp}(\mathbf{x})\subseteq \mathsf{S}} l(\mathbf{x}) \tag{2}$$

が弱劣モジュラ性 [6] を持つことが明らかになっており，貪欲法のような集合関数最大化に対する手法も，問題 (1) に対する有望なアプローチの一つとして注目されている．

本稿ではまず，上記の集合関数 $F(\cdot)$ が弱劣モジュラ性だけでなく，弱優モジュラ性 [4] も有していることを確認する．このことを用いて，問題 (1) が制限強凸性・制限平滑性の意味で良条件な場合，$F(\cdot)$ をモジュラ関数で近似して最大化するだけでも良い解が得られることや，問題 (1) に対する fixed-parameter tractable (FPT) な乱択近似アルゴリズムを構成できることを示す．次に，問題 (1) の制約が単調劣モジュラ関数 $G : 2^{[d]} \to \mathbb{R}$ と $b > 0$ によって $\mathcal{F} = \{\mathsf{S} \subseteq [d] \mid G(\mathsf{S}) \le b\}$ のように書ける場合を考える．この問題

は ℓ_0-制約付き最小化問題を含むクラスを成している．このようなサポートが劣モジュラ的な構造を持つ最適化問題に対しては，凸最適化としての定式化に基づく研究が行われている [1, 2]．一方で，上記のように非凸最適化問題として定式化される場合についての研究は少なかった．本項では，$G(\cdot)$ が弱優モジュラ性を持っている場合に注目し，貪欲法や IHT の理論保証を与える．

2 準備

本項では，$[d]$ の部分集合は S や T のような大文字サンセリフ体で記述し，$[d]$ の要素を書くのに j を用いる．また，部分集合 $\{j\} \subseteq [d]$ をしばしば単に j と略記する．$[d]$ 上の集合関数は F や G のように大文字で書く．与えられた集合関数 $F : 2^{[d]} \to \mathbb{R}$ と任意の $\mathsf{S}, \mathsf{T} \subseteq [d]$ に対し，$F(\mathsf{T} \mid \mathsf{S}) \coloneqq F(\mathsf{S} \cup \mathsf{T}) - F(\mathsf{S})$ と定義する．本項では，単調な集合関数のみを考える．すなわち，任意の $\mathsf{S}, \mathsf{T} \subseteq [d]$ に対し，$F(\mathsf{T} \mid \mathsf{S}) \geq 0$ が成り立つもののみを考える．任意の $\mathsf{S} \subseteq \mathsf{T}$ と $j \notin \mathsf{T}$ に対し，$F(j \mid \mathsf{S}) \geq F(j \mid \mathsf{T})$ が成り立つとき $F(\cdot)$ は劣モジュラであると言い，$F(j \mid \mathsf{S}) \leq F(j \mid \mathsf{T})$ が成り立つとき $F(\cdot)$ は優モジュラであると言う．

次に，式 (2) で定義される集合関数 $F(\cdot)$ について説明する．$F(\cdot)$ は単調であり，$F(\emptyset) = 0$ を満たす．以下では，任意の $\mathsf{S} \in \mathcal{F}$ に対し関数値 $F(\mathsf{S})$ の計算は d についての多項式時間（以下 $\mathrm{poly}(d)$ 時間と記す）で行えると仮定する．

本項では，ベクトルは $\mathbf{x}$ や $\mathbf{y}$ のような太字の小文字で記し，ゼロベクトルは単に 0 と書く．任意の $\mathsf{S} \subseteq [d]$ と $\mathbf{x} \in \mathbb{R}^{[d]}$ に対し，$\mathbf{x}_{\mathsf{S}} \in \mathbb{R}^{\mathsf{S}}$ は S に含まれる添字に対応する成分のみからなる $\mathbf{x}$ の部分ベクトルを表すものとする．

弱劣モジュラ性と弱優モジュラ性　単調集合関数 $F : 2^{[d]} \to \mathbb{R}$ の弱劣モジュラ性・弱優モジュラ性はそれぞれ，弱劣モジュラ比・弱優モジュラ比と呼ばれるパラメータを用いて定義される．$\mathsf{U} \subseteq [d]$ と $k > 0$ をそれぞれ任意の固定された部分集合と正整数とする．このとき，$F(\cdot)$ の弱劣モジュラ比 $\gamma_{\mathsf{U},k}$ と弱優モジュラ比 $\beta_{\mathsf{U},k}$ は以下のように定義される．

$$\gamma_{\mathsf{U},k} \coloneqq \min_{\substack{\mathsf{L},\mathsf{S}\,:\,\mathsf{L}\cap\mathsf{S}=\emptyset,\\ \mathsf{L}\subseteq\mathsf{U},\,|\mathsf{S}|\leq k}} \frac{\sum_{j\in\mathsf{S}} F(j \mid \mathsf{L})}{F(\mathsf{S} \mid \mathsf{L})}, \qquad \beta_{\mathsf{U},k} \coloneqq \max_{\substack{\mathsf{L},\mathsf{S}\,:\,\mathsf{L}\cap\mathsf{S}=\emptyset,\\ \mathsf{L}\subseteq\mathsf{U},\,|\mathsf{S}|\leq k}} \frac{\sum_{j\in\mathsf{S}} F(j \mid \mathsf{L})}{F(\mathsf{S} \mid \mathsf{L})}.$$

ただし $0/0 = 1$ とみなす．また，$\gamma_{k',k} \coloneqq \min_{|\mathsf{U}|\leq k'} \gamma_{\mathsf{U},k}$，$\beta_{k',k} \coloneqq \max_{|\mathsf{U}|\leq k'} \beta_{\mathsf{U},k}$ と定義する．単調性から，任意の U, k に対して $\gamma_{\mathsf{U},k} \in [0,1]$ および $\beta_{\mathsf{U},k} \in [1,k]$ が成り立つ．また，定義から任意の $\mathsf{U}' \subseteq \mathsf{U}$ と $k' \leq k$ について $\gamma_{\mathsf{U},k} \leq \gamma_{\mathsf{U}',k'}$ および $\beta_{\mathsf{U},k} \geq \beta_{\mathsf{U}',k'}$ が成り立つ．$\gamma_{d,d} = 1$ ならば $F(\cdot)$ は劣モジュラであり，$\beta_{d,d} = 1$ ならば $F(\cdot)$ は優モジュラである．$\gamma_{d,d} = \beta_{d,d} = 1$ ならば $F(\cdot)$ はモジュラ関数となる．

制限強凸性と制限平滑性　ある固定された $\Omega \subseteq \mathbb{R}^{[d]} \times \mathbb{R}^{[d]}$ が与えられたもとで，連続的微分可能関数 $l : \mathbb{R}^{[d]} \to \mathbb{R}$ が μ_Ω-制限強凸，ν_Ω-制限平滑であるとは

$$\frac{\mu_\Omega}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 \leq l(\mathbf{y}) - l(\mathbf{x}) - \langle \nabla l(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \leq \frac{\nu_\Omega}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 \tag{3}$$

が全ての $(\mathbf{x}, \mathbf{y}) \in \Omega$ に対して成り立つことを言う．正実数 μ_Ω，ν_Ω はそれぞれ制限強凸定数，制限平滑定数と呼ばれる．ここで，$\Omega' \subseteq \Omega$ であるならば，$\mu_{\Omega'} \geq \mu_\Omega$ および $\nu_{\Omega'} \leq \nu_\Omega$ が成り立つように制限強凸定数・制限平滑定数を定めることができる点に注意する．以下では，$\Omega = \Omega_{k_1,k_2} \coloneqq \{(\mathbf{x}, \mathbf{y}) \mid \|\mathbf{x}\|_0 \leq k_1, \|\mathbf{y}\|_0 \leq k_1, \|\mathbf{x} - \mathbf{y}\|_0 \leq k_2\}$ について式 (3) が成り立つとき，$l(\cdot)$ は μ_{k_1,k_2}-制限強凸および ν_{k_1,k_2}-制限平滑であると言うことにする．また，$\mu_k \coloneqq \mu_{k,k}$，$\nu_k \coloneqq \nu_{k,k}$ と定義する．

3　制限強凸性・制限平滑性と弱劣モジュラ性・弱優モジュラ性の関係

集合関数 $F(\cdot)$ が式 (2) で定義されているとする．このとき，[8] にあるように，$F(\cdot)$ の劣モジュラ比 $\gamma_{\mathsf{U},k}$ は $l(\cdot)$ の制限強凸定数と制限平滑定数の比を用いて下から抑えることができる（下記の定理にも詳細を記す）．また，関数が定義域全域において強凸かつ平滑な場合は，強凸定数と平滑定数の比によって $F(\cdot)$ の優モジュラ比の上界が得られることが知られている [4]．本研究ではまず，$l(\cdot)$ が定義域全体における強凸性や平滑性を持たない場合であっても，適切な領域上での制限強凸性や制限平滑性を持っているならば，$F(\cdot)$ の優モジュラ比 $\beta_{\mathsf{U},k}$ の上界が得られることを示した．

定理 1 (劣モジュラ比の下界は [8] からの引用)**.** 任意の $\mathsf{U} \subseteq [d]$ と $k \in \mathbb{Z}_{>0}$ に対し，式 (2) で定義される $F(\cdot)$ の劣モジュラ比 $\gamma_{\mathsf{U},k}$ の下界と優モジュラ比 $\beta_{\mathsf{U},k}$ の上界はそれぞれ，$l(\cdot)$ の制限強凸定数・制限平滑定数を用いて次のように表せる．

$$\gamma_{\mathsf{U},k} \geq \frac{\mu_{|\mathsf{U}|+k}}{\nu_{|\mathsf{U}|+1,1}} \geq \frac{\mu_{|\mathsf{U}|+k}}{\nu_{|\mathsf{U}|+k}}, \qquad \beta_{\mathsf{U},k} \leq \frac{\nu_{|\mathsf{U}|+k,k}}{\mu_{|\mathsf{U}|+1}} \leq \frac{\nu_{|\mathsf{U}|+k}}{\mu_{|\mathsf{U}|+k}}.$$

$\kappa_{|\mathsf{U}|+k} := \nu_{|\mathsf{U}|+k}/\mu_{|\mathsf{U}|+k} \geq 1$ は制限条件数と呼ばれる [11]．端的に言えば，目的関数の制限条件数が小さいほど，問題 (1) は扱いやすくなる．一方で，劣モジュラ比・優モジュラ比は集合関数 $F(\cdot)$ のモジュラ関数への近さを表しており，$\kappa_{|\mathsf{U}|+k} \approx 1$ ならば $F(\mathsf{S} \mid \mathsf{L}) \approx \sum_{j \in \mathsf{S}} F(j \mid \mathsf{L})$ となる．これらの事実は，制限条件数の観点から見た問題 (1) の扱いやすさと，集合関数最大化問題の扱いやすさのと繋がりを示唆している．このことから，制限条件数 $\kappa_{|\mathsf{U}|+k}$ が小さいならば，$F(\cdot)$ をモジュラ関数によって近似し，その結果得られた関数を最大化することで，ある程度良い解を得ることができる．このような手法はスパース最適化の文脈では oblivious support selection [8] と呼ばれており，定理 1 を用いれば，この手法に対する以下の保証が得られる．

系 1.a. $F(\cdot)$ は式 (2) で定義されているとし，$\mathsf{S} = \mathrm{argmax}_{\mathsf{T} \in \mathcal{F}} \sum_{j \in \mathsf{T}} F(j)$，$\mathsf{S}^* = \mathrm{supp}(\mathbf{x}^*)$ とする．ただし $\mathbf{x}^*$ は問題 (1) の最適解である．また，$k = \max\{|\mathsf{S}|, |\mathsf{S}^*|\}$ とする．このとき，

$$\begin{aligned} F(\mathsf{S}) &\geq \frac{1}{\beta_{\emptyset,|\mathsf{S}|}} \sum_{j \in \mathsf{S}} F(j) \geq \frac{1}{\beta_{\emptyset,|\mathsf{S}|}} \sum_{j \in \mathsf{S}^*} F(j) \\ &\geq \frac{\gamma_{\emptyset,|\mathsf{S}^*|}}{\beta_{\emptyset,|\mathsf{S}|}} F(\mathsf{S}^*) \geq \frac{\mu_1}{\nu_{|\mathsf{S}|}} \frac{\mu_{|\mathsf{S}^*|}}{\nu_1} F(\mathsf{S}^*) \geq \frac{1}{\kappa_1 \kappa_k} F(\mathsf{S}^*) \end{aligned}$$

が成り立つ．特に，ある $\epsilon \geq 0$ に対して $\kappa_1 \kappa_k \leq 1 + \epsilon$ が成り立つならば，$\mathbf{x} = \mathrm{argmin}_{\mathrm{supp}(\mathbf{x}') \subseteq \mathsf{S}} l(\mathbf{x}')$ は $l(\mathbf{x}) \leq l(\mathbf{x}^*) + \epsilon(l(0) - l(\mathbf{x}))$ を満たす．

この結果は様々な制約に対して適用することができる．例えば，$([d], \mathcal{F})$ がマトロイドであれば，$F(\cdot)$ を近似するモジュラ関数は貪欲法によって最大化できる [14]．

3.1　ℓ_0-制約付き最小化問題に対する乱択 FPT 近似アルゴリズム

Algorithm 1 乱択 FPT 近似アルゴリズム

```
SingleRun() を T 回実行し，その中で目的関数値が最大となる解を出力する．
function SingleRun()
    S_0 ← ∅
    for i = 1, ..., k do
        j ∈ [d]\S_{i−1} を F(j | S_{i−1}) に比例する確率で選ぶ．
        S_i ← S_{i−1} ∪ {j}
    return S_k
```

本節では，以下の ℓ_0-制約付き最小化問題に対する乱択 FPT 近似アルゴリズムについて考える．

$$\underset{\mathbf{x}\in\mathbb{R}^{[d]}}{\text{minimize}}\ \ l(\mathbf{x}) \quad \text{subject to}\ \ \|\mathbf{x}\|_0 \le k. \tag{4}$$

この問題は一般には NP 困難であり [16]，既存の多項式時間解法の理論保証 [13, 17] は，k の値が十分大きい（場合によっては最適解 $\mathbf{x}^*$ の非ゼロ成分数より大きい）ことを仮定して導出されることがほとんどである．k に関する仮定を置かずに任意の近似精度を達成する素朴な方法としては，全通りのサポートを試す方法があるが，その計算量は $\Omega(d^k)$ となってしまう．以下では，parametrized complexity [5] の考え方に基づいて，d についての計算量が $\mathrm{poly}(d)$ となるような手法を構築することを考える．具体的には，d 以外の入力（の一部）$\mathbf{p}$ を固定パラメータ (fixed parameter) とみなし，計算量が $g(\mathbf{p})\cdot\mathrm{poly}(d)$ となるような fixed-parameter-tractable アルゴリズム（FPT アルゴリズム）を設計する．ただし，$g(\cdot)$ は $\mathbf{p}$ についての計算可能な関数である．

アルゴリズムのアイデアは，p-分割可能な単調劣モジュラ関数最大化に対する乱択 FPT 近似アルゴリズム [18] のものに基づいている．まずこのアリゴリズムに対する解析を，以下の弱劣モジュラかつ弱優モジュラな集合関数の最大化問題の場合に拡張する．

$$\underset{\mathsf{S}\subseteq[d]}{\text{minimize}}\ \ F(\mathsf{S}) \quad \text{subject to}\ \ |\mathsf{S}| \le k. \tag{5}$$

この問題に対し，Algorithm 1 を適用する．Algorithm 1 は乱択貪欲法 `SingleRun()` を T 回実行し，得られた解の中で最良の解を出力する．このアルゴリズムに対し，以下の近似保証が得られる．

定理 2 ([18] の拡張)**.** $F(\cdot)$ が劣モジュラ比 $\gamma_{k,k}$ と優モジュラ比 $\beta_{k,d}$ を持っているとし，S^* を問題 (5) に対する最適解とする．任意の $\epsilon > 0$ に対し，*Algorithm 1* の T が

$$T \ge \left\lceil \left(\frac{\beta_{k,d}}{\gamma_{k,k}}\cdot\frac{1+\epsilon}{\epsilon}\right)^k \log\delta^{-1}\right\rceil$$

を満たすならば，出力された解 S は $F(\mathsf{S}) \ge (1+\epsilon)F(\mathsf{S}^*) - \epsilon F([d])$ を少なくとも確率 $1-\delta$ で満たす．

さらに，Algorithm 1 を実行して得られた解を S とし，$\mathbf{x} = \mathrm{argmin}_{\mathrm{supp}(\mathbf{x}')\subseteq\mathsf{S}}\, l(\mathbf{x}')$ を問題 (4) に対する解としたとき，定理 1, 2 から以下の結果が得られる．

系 2.a. $\mathbf{x}^*$ を問題 (4) に対する最適解とし，$l_{\min} \coloneqq \min_{\mathbf{x}\in\mathbb{R}^{[d]}} l(\mathbf{x})$ とする．また，$\mathbf{p} \coloneqq (k, \mu_{2k}, \mu_{k+1}, \nu_{k+1,1}, \nu_d, \epsilon, \delta)$ を固定パラメータとみなす．このとき，*Algorithm 1* を用いて $\|\mathbf{x}\|_0 \le k$ および $l(\mathbf{x}) \le l(\mathbf{x}^*) + \epsilon(l(\mathbf{x}^*) - l_{\min})$ を満たす解 $\mathbf{x}$ を確率 $1-\delta$ で計算できる．ただし，*Algorithm 1* は

$$O\left(\left\lceil \left(\frac{\nu_{k+1,1}}{\mu_{2k}}\cdot\frac{\nu_d}{\mu_{k+1}}\cdot\frac{1+\epsilon}{\epsilon}\right)^k \log\delta^{-1}\right\rceil k\cdot d\right) \eqqcolon g(\mathbf{p})\cdot d$$

回 $F(\cdot)$ の値を評価する．

$F(\cdot)$ の関数値評価は $\mathrm{poly}(d)$ 時間で行えることを仮定しているため，Algorithm 1 は $O(\epsilon)$ 誤差近似解を高確率で計算できる $g(\mathbf{p})\cdot d\times\mathrm{poly}(d) = g(\mathbf{p})\cdot\mathrm{poly}(d)$ 時間のアルゴリズムとなっている．

4　単調劣モジュラ制約付き最小化問題

集合関数 $G: 2^{[d]} \to \mathbb{R}$ を $G(\emptyset) = 0$ を満たす単調劣モジュラ関数とし，$G(\mathrm{supp}(\mathbf{x}))$ が所定の上限値以下となるような解 $\mathbf{x}\in\mathbb{R}^{[d]}$ 求める最小化問題を考える．このような，サポートについて劣モジュラな構造を持った問題は，機械学習の様々な場面で現れる [1]．以下では $\rho \coloneqq \max_{j\in[d]} G(j)$ とする．ここで，$G(j)$ が上限値より大きい j を含む解は単調性から常に実行不可能であるため，ρ は制約の上限値以下であるとみ

Algorithm 2 貪欲法

$\mathsf{U} \leftarrow [d]$, $\mathsf{S} \leftarrow \emptyset$
while $\mathsf{U} \neq \emptyset$ **do**
  $j \leftarrow \mathrm{argmax}_{j' \in \mathsf{U}} \frac{F(j'|\mathsf{S})}{G(j'|\mathsf{S})}$
  if $G(\mathsf{S} \cup \{j\}) \leq c + \rho$ **then**
    $\mathsf{S} \leftarrow \mathsf{S} \cup \{j\}$
  $\mathsf{U} \leftarrow \mathsf{U} \backslash \{j\}$
return S

なしてよい．これを考慮して，制約の上限値をある $c \geq 0$ を用いて $c + \rho$ と書く．このとき，解くべき問題は以下のように定式化できる．

$$\underset{\mathbf{x} \in \mathbb{R}^{[d]}}{\mathrm{minimize}} \ \ l(\mathbf{x}) \quad \text{subject to} \ \ G(\mathrm{supp}(\mathbf{x})) \leq c + \rho. \tag{6}$$

ℓ_0-制約付き最小化の文脈では，真のスパース解の非ゼロ成分数よりも多くの非ゼロ成分を取ることを許して，アルゴリズムの理論保証について議論することが多い [13, 17]．このことを踏まえて，以下では $c^* \leq c + \rho$ を満たす c^* を固定し，$\mathbf{x}^* := \mathrm{argmin}_{G(\mathrm{supp}(\mathbf{x})) \leq c^*} l(\mathbf{x})$ を目的の解と見なしてアルゴリズムの解析を行う．また，$k^* := \|\mathbf{x}^*\|_0$ と定義する．以下では次の条件が成り立っていると仮定する．

仮定 1. 全ての $j \in [d]$ は $G(j) \leq c^*$ を満たす．

この仮定を破る j は $\mathrm{supp}(\mathbf{x}^*)$ に含まれないため，あらかじめコストの大きすぎる j を除外したと考えれば，この仮定は自然である．これらの条件のもとで，問題 (6) に対する貪欲法と IHT について考える．

4.1 貪欲法

式 (2) で定義された $F(\cdot)$ を用いて，次のような弱劣モジュラ関数の最大化問題を考える．

$$\underset{\mathsf{S} \subseteq [d]}{\mathrm{maximize}} \ \ F(\mathsf{S}) \quad \text{subject to} \ \ G(\mathsf{S}) \leq c + \rho.$$

この問題に対して Algorithm 2 を適用する．Algorithm 2 はナップサック制約下の単調劣モジュラ関数最大化問題に対する貪欲法 [15, 19] を，単調劣モジュラ関数制約の場合に拡張したものとして見ることができる（ただし，一部元の手法をやや単純化している）．この貪欲法について以下の近似保証が成り立つ．

定理 3. S をアルゴリズムの出力とし $\mathsf{S}^* := \mathrm{supp}(\mathbf{x}^*)$ と定義する．$F(\cdot)$ が劣モジュラ比 γ_{S,k^*} を持ち，$G(\cdot)$ が優モジュラ比 $\beta_{\emptyset,k^*}$ を持つならば，

$$F(\mathsf{S}) \geq \left(1 - \exp\left(-\frac{\gamma_{\mathsf{S},k^*}}{\beta_{\emptyset,k^*}} \cdot \frac{c}{c^*}\right)\right) F(\mathsf{S}^*)$$

が成り立つ．

この結果と定理 1 より，問題 (6) に対して次の結果が得られる．

系 3.a. 定理 3 と同様の条件が成り立っているとし，$\mathbf{x} := \mathrm{argmin}_{\mathrm{supp}(\mathbf{x}') \subseteq \mathsf{S}} l(\mathbf{x}')$，$k := |\mathsf{S}|$ とする．$l(\cdot)$ が μ_{k+k^*}-制限強凸かつ $\nu_{k+1,1}$-制限平滑ならば，$l(\mathbf{x}) \leq l(\mathbf{x}^*) + \exp\left(-\frac{\mu_{k+k^*}}{\nu_{k+1,1}} \cdot \frac{c}{c^* \beta_{\emptyset,k^*}}\right)(l(0) - l(\mathbf{x}^*))$ が成り立つ．特に，任意の $\epsilon > 0$ を固定したとき，

$$c \geq c^* \beta_{\emptyset,k^*} \frac{\nu_{k+1,1}}{\mu_{k+k^*}} \log \frac{l(0) - l(\mathbf{x}^*)}{\epsilon}$$

が成り立つならば，$l(\mathbf{x}) \leq l(\mathbf{x}^*) + \epsilon$ となる．

この結果は，ℓ_0-制約付き最小化問題に対する貪欲法の結果 [17] を単調劣モジュラ関数制約の場合に拡張したものとして見ることができる（ただし，制約の上限値が $c+\rho$ の形で与えらているため，完全な一般化ではない）．

4.2 IHT

Algorithm 3 IHT

1. 初期解 $\mathbf{x}_0 \in \mathbb{R}^{[d]}$ を定める．
2. **for** $t = 0, 1, \ldots, T$ **do**
3. $\quad \mathbf{g}_t \leftarrow \mathbf{x}_t - \eta \nabla l(\mathbf{x}_t)$
4. $\quad \mathbf{x}_{t+1} \leftarrow \mathcal{P}_c(\mathbf{g}_t)$
5. **return** $\mathbf{x}_T$

次に，射影勾配法に基づく手法 (IHT) を問題 (6) に適用することを考える．ℓ_0-制約付き最小化問題に対する IHT [13] と同様に，Algorithm 3 は勾配降下と実行可能領域への射影を繰り返して解を更新する．ただし，問題 (6) の制約は単調劣モジュラ関数 $G(\cdot)$ によって与えられているため，それを考慮した射影ステップ（Algorithm 3 の $\mathcal{P}_c(\mathbf{g}_t)$）を考える必要がある．ここでは，この射影ステップを貪欲法によって行うことを考える．具体的には，目的関数 $F(\mathsf{S}) = \|(\mathbf{g}_t)_{\mathsf{S}}\|_2^2$ と制約 $G(\mathsf{S}) \leq c + \rho$ に対して Algorithm 2 を実行して得られた解 S を用い，$j \in \mathsf{S}$ ならば $(\mathbf{x}_{t+1})_j$ を $(\mathbf{g}_t)_j$ とし，$j \notin \mathsf{S}$ ならば 0 とする．定理 3 からこの射影ステップの近似精度を評価できるため，IHT について以下の理論保証が得られる．

定理 4. $k \coloneqq \max_{t:0\leq t \leq T} \|\mathbf{x}_t\|_0$，$\omega \coloneqq \max_{t:0 \leq t \leq T} \|\mathbf{g}_t\|_2$ とする．$l(\cdot)$ は連続的二階微分可能であり，μ_{2k+k^*}-制限強凸，ν_{2k+k^*}-制限平滑 であると仮定する．また，$G(\cdot)$ は優モジュラ比 $\beta_{\emptyset,k^*}$ を持つと仮定する．さらに $\eta = \frac{1}{\nu_{2k+k^*}}$ とする．この時，

$$c \geq 4c^* \beta_{\emptyset,k^*} \left(\frac{\nu_{2k+k^*}}{\mu_{2k+k^*}}\right)^2 + 2c^*\beta_{\emptyset,k^*} \log\left(\frac{\omega}{2\epsilon}\right) + \rho$$

であるならば，

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2 \leq \left(1 - \frac{1}{2} \cdot \frac{\mu_{2k+k^*}}{\nu_{2k+k^*}}\right) \|\mathbf{x}_t - \mathbf{x}^*\|_2 + \zeta + \frac{\mu_{2k+k^*}}{\nu_{2k+k^*}} \cdot \epsilon$$

が成立する．ただし，$\zeta \coloneqq \frac{1}{\nu_{2k+k^*}} \left(1 + \frac{1}{2} \cdot \frac{\mu_{2k+k^*}}{\nu_{2k+k^*}}\right) \max_{\mathsf{S} \subseteq [d]} \{\|\nabla l(\mathbf{x}^*)_{\mathsf{S}}\|_2 \mid G(\mathsf{S}) \leq c + \rho \ and\ |\mathsf{S}| \leq k\}$ である．特に

$$T \geq 2 \cdot \frac{\nu_{2k+k^*}}{\mu_{2k+k^*}} \log \frac{\|\mathbf{x}_0 - \mathbf{x}^*\|_2}{\epsilon}$$

を満たす T について，$\|\mathbf{x}_T - \mathbf{x}^*\|_2 \leq 3\epsilon + 2\zeta \cdot \frac{\nu_{2k+k^*}}{\mu_{2k+k^*}}$ が成り立つ．

上記の定理は，c が十分大きな値をとるならば，十分に反復を繰り返すことで，$\mathbf{x}^*$ を $O(\zeta \cdot \frac{\nu_{2k+k^*}}{\mu_{2k+k^*}})$ 誤差の精度で復元できるということを意味している．特に，$\mathbf{x}_{\min} \coloneqq \mathrm{argmin}_{\mathbf{x} \in \mathbb{R}^{[d]}} l(\mathbf{x})$ が $G(\mathrm{supp}(\mathbf{x}_{\min})) \leq c^*$ を満たすならば $\zeta = 0$ が成り立つ．

5 まとめ

本稿では，まず非凸制約付き最小化問題に由来する集合関数最大化問題を考え，元の問題の目的関数が制限強凸性・制限平滑性を持つならば，得られた集合関数は弱劣モジュラ性・弱優モジュラ性を持つことを確認した．この性質を用いて，ℓ_0-制約付き最小化問題に対する乱択 FPT 近似アルゴリズムを与えた．さらに，制約がサポートついての単調劣モジュラ関数を用いて表現されるような最小化問題を考え，その問題に対する貪欲法と IHT の理論保証を与えた．

6 謝辞

この研究は京都大学数理解析研究所の共同利用・共同研究による成果である．

参考文献

[1] F. Bach. Structured sparsity-inducing norms through submodular functions. In *Advances in Neural Information Processing Systems 23*, pages 118–126. Curran Associates, Inc., 2010.

[2] F. Bach. Learning with submodular functions: A convex optimization perspective. *Foundations and Trends® in Machine Learning*, 6(2-3):145–373, 2013.

[3] T. Blumensath and M. E. Davies. Iterative hard thresholding for compressed sensing. *Appl. Comput. Harmon. Anal.*, 27(3):265–274, 2009.

[4] I. Bogunovic, J. Zhao, and V. Cevher. Robust maximization of non-submodular objectives. In *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 890–899. PMLR, 2018.

[5] M. Cygan, F. V. Fomin, L. Kowalik, D. Lokshtanov, D. Marx, M. Pilipczuk, M. Pilipczuk, and S. Saurabh. *Parameterized Algorithms*. Springer Publishing Company, Incorporated, 1st edition, 2015.

[6] A. Das and D. Kempe. Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection. In *Proceedings of the 28th International Conference on Machine Learning*, pages 1057–1064. ACM, 2011.

[7] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[8] E. R. Elenberg, R. Khanna, A. G. Dimakis, and S. Negahban. Restricted strong convexity implies weak submodularity. *Ann. Statist.*, 46(6B):3539–3568, 2018.

[9] S. Foucart. Hard thresholding pursuit: An algorithm for compressive sensing. *SIAM J. Optim.*, 49(6):2543–2563, 2011.

[10] C. Hegde, P. Indyk, and L. Schmidt. A nearly-linear time framework for graph-structured sparsity. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, pages 928–937. PMLR, 2015.

[11] P. Jain and P. Kar. Non-convex optimization for machine learning. *Foundations and Trends® in Machine Learning*, 10(3-4):142–336, 2017.

[12] P. Jain, N. Rao, and I. S. Dhillon. Structured sparse regression via greedy hard thresholding. In *Advances in Neural Information Processing Systems 29*, pages 1516–1524. Curran Associates, Inc., 2016.

[13] P. Jain, A. Tewari, and P. Kar. On iterative hard thresholding methods for high-dimensional M-estimation. In *Advances in Neural Information Processing Systems 27*, pages 685–693. Curran Associates, Inc., 2014.

[14] B. Korte and J. Vygen. *Combinatorial Optimization*, volume 2. Springer, 2012.

[15] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 420–429. ACM, 2007.

[16] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM J. Optim.*, 24(2):227–234, 1995.

[17] S. Shalev-Shwartz, N. Srebro, and T. Zhang. Trading accuracy for sparsity in optimization problems with sparsity constraints. *SIAM J. Optim.*, 20(6):2807–2832, 2010.

[18] P. Skowron. FPT approximation schemes for maximizing submodular functions. *Inform. Comput.*, 257:65 – 78, 2017.

[19] M. Sviridenko. A note on maximizing a submodular set function subject to a knapsack constraint. *Oper. Res. Lett.*, 32(1):41–43, 2004.

[20] X. Yuan, P. Li, and T. Zhang. Gradient hard thresholding pursuit for sparsity-constrained optimization. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32, pages 127–135. PMLR, 2014.